DOI: 10.1534/genetics.105.041780

Origins of Host-Specific Populations of the Blast Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of Pandemic Clones on Rice and Weeds of Rice

Brett C. Couch,*[,1] Isabelle Fudal,[†] Marc-Henri Lebrun,[†] Didier Tharreau,[‡] Barbara Valent,[§] Pham van Kim, Jean-Loup Nottéghem[‡] and Linda M. Kohn*[,2]**

**Department of Botany, University of Toronto, Mississauga, Ontario L5L 1C6, Canada, [†]Centre National de la Recherche Scientifique, Bayer Cropscience, Lyon 69263, France, [‡]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR BGPI, Montpellier 34398, France, [§]Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506 and **Department of Plant Pathology, Can Tho University, Can Tho, Vietnam*

Manuscript received February 8, 2005
Accepted for publication March 2, 2005

ABSTRACT

Rice, as a widely and intensively cultivated crop, should be a target for parasite host shifts and a source for shifts to co-occurring weeds. *Magnaporthe oryzae*, of the *M. grisea* species complex, is the most important fungal pathogen of rice, with a high degree of host specificity. On the basis of 10 loci from six of its seven linkage groups, 37 multilocus haplotypes among 497 isolates of *M. oryzae* from rice and other grasses were identified. Phylogenetic relationships among isolates from rice (*Oryza sativa*), millet (*Setaria* spp.), cutgrass (*Leersia hexandra*), and torpedo grass (*Panicum repens*) were predominantly tree like, consistent with a lack of recombination, but from other hosts were reticulate, consistent with recombination. The single origin of rice-infecting *M. oryzae* followed a host shift from a Setaria millet and was closely followed by additional shifts to weeds of rice, cutgrass, and torpedo grass. Two independent estimators of divergence time indicate that these host shifts predate the Green Revolution and could be associated with rice domestication. The rice-infecting lineage is characterized by high copy number of the transposable element MGR586 (Pot3) and, except in two haplotypes, by a loss of *AVR-Co39*. Both mating types have been retained in ancestral, well-distributed rice-infecting haplotypes 10 (mainly temperate) and 14 (mainly tropical), but only one mating type was recovered from several derived, geographically restricted haplotypes. There is evidence of a common origin of both *ACE1* virulence genotypes in haplotype 14. Host-haplotype association is evidenced by low pathogenicity on hosts associated with other haplotypes.

WIDELY distributed, high-density populations, such as those of cultivated crops or humans, are large targets for host shifts by parasites. After crossing host species boundaries, successful shifts culminate with specialization and genetic divergence on the new host (Antonovics *et al.* 2002). If the parasite crosses a species barrier only once, virulence determinants would derive from this common origin. In contrast, if the target host species is invaded in multiple, independent host shifts, virulence could have multiple, fundamentally different determinants among host populations (MacLeod *et al.* 2001).

After successful establishment on a new host, especially one that is abundant and relatively genetically uniform, expansion of pandemic clones of the pathogen commonly follows (Tibayrenc *et al.* 1991; Maynard Smith *et al.* 1993; Brasier and Buck 2001). While we would expect that virulence determinants in a pandemic clone would remain associated and modified only through mutation, any genetic exchange with recombination would shuffle these determinants into novel combinations across the parasite population. Within large pathogen populations, recombination would increase the rate at which multiple virulence factors could combine to generate "super parasite" genotypes (Crow and Kimura 1970).

The rice blast pathosystem is a model system not only for plant-pathogen interactions (Valent 1990; Talbot 2003), but also for host shifts on a relatively contemporary timescale among grass species. We are mainly interested in the host shift to rice. While the genetic basis of the host-pathogen interaction fits the criteria for coevolution, the time frame does not appear to be consistent with cospeciation. The earliest report of a blast-like disease of rice (*Oryza sativa*) in the literature is 1637 in China and, in the following years, in Japan (Ou 1987). The degree of phylogenetic divergence among grass hosts is large compared to the phylogenetic divergence among

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY944075–AY944131 and AY944359–AY944405.

[1]*Present address:* Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55108.

[2]*Corresponding author:* Department of Botany, University of Toronto, 3359 Mississauga Rd. N, Mississauga, Ontario L5L 1C6, Canada. E-mail: kohn@utm.utoronto.ca

pathogen genotypes from different hosts (Couch and Kohn 2002). Within the grasses are two major clades estimated to have originated at least 55 million years ago, comprising ~10,000 species (Crepet and Feldman 1991; Grass Phylogeny Working Group 2001; Bremer 2002). Genotypes of the pathogen species are highly host specific, yet in sum infect grasses from both major clades, suggesting a rapid, recent radiation on grasses. Given this time scale, we see domestication as the best opportunity for initial host shifts to rice, as large populations of cultivated rice are likely targets for pathogenic migrants from other hosts.

There is now good archeological evidence and more consensus among experts in support of an early domestication of *O. sativa* from its wild relative *O. rufipogon* in the middle Yangtze valley in China ~7000 years before present (BP) (Crawford and Shen 1998; Higman and Lu 1998; Vaughan and Morishima 2003). Chang (1976) presented evidence for early rice cultivation in the Indian Himalayas 2450–3950 BP and in China 5230 BP, proposing multiple, independent, concurrent domestication events across the region stretching from the foothills of the Indian Himalayas to southwestern and southern China. Separate domestication of the two varieties *indica* and *japonica* (Second 1982) and, alternatively, domestication of *indicas* from wild rice and derivation of *japonicas* from *indicas* (Oka 1988) have been proposed, but the evidence is equivocal (Lu *et al.* 2002; Cheng *et al.* 2003; Yamanaka *et al.* 2003).

The expansion of rice cultivation to northern China, Korea, and Japan brought two domesticated hosts of the blast pathogen, rice, and Italian millet (*Setaria italica*) into close contact in crop complexes between 7000 BP in China and 3000 BP in Korea and Japan (Crawford 1992; Higman and Lu 1998). Domesticated rice spread throughout Asia, the Mediterranean (~2300 BP), Madagascar (~1500 BP), Southern Europe (500–600 BP), and, more recently, as a result of European colonization, to Africa and North and South America (~300–500 BP) (Oka 1988). Introduction of *O. sativa* to Africa also brought it in contact with the more recently domesticated (~3500 BP), closely related, West African domesticated species, *O. glaberrima* (Chang 1976). Intercontinental dispersal of rice genotypes continues by commercial trade, official distribution of new rice cultivars from centers of plant breeding, and unofficial (in the pocket) exchange. Dispersal of rice can also disperse rice blast (Long *et al.* 2001). Even if rare, seed transmission has probably introduced, and reintroduced, rice blast to several countries.

The Green Revolution of the 1960s popularized the first high-yielding, semidwarf rice cultivars, abruptly changing the diversity and frequency of rice genotypes throughout Asia (Khush 2001). The responsiveness of these varieties to nitrogen fertilizer led to increased inputs that increased the susceptibility of rice to blast and set the stage for the invasion of novel blast genotypes into modern rice (Ou 1987). Existing blast pathogen populations may have undergone a bottleneck with reduction in the diversity of rice genotypes. Monoculture conditions would have facilitated the expansion of pandemic blast clones. While the Green Revolution may have caused a shift in blast populations, the "revolution" was incremental. It is likely that some blast genotypes predating the Green Revolution, or their descendants, still exist.

At either end of the timescale for invasion by the blast pathogen—7000 or 35 years ago—models would predict that the high frequency and density of rice crops, compared to other hosts, would favor specialization for rice among invading parasite genotypes (Fry 1996; Whitlock 1996). Host shifts are the most plausible explanation for the epidemics of blast in the mid-1980s on wheat (*Triticum aestivum*) in Brazil and in the early 1990s on perennial ryegrass (*Lolium perenne*) in Pennsylvania (Landschoot and Hoyland 1992; Oh *et al.* 2002).

The biological characteristics of the blast pathogen should enable invasions leading to host shifts. Rice blast, and gray leaf spot of grasses, is caused by *Magnaporthe oryzae*, a taxon recently segregated from *M. grisea*, a species complex of haploid ascomycetous fungi (Couch and Kohn 2002). Hosts of the *M. grisea-M. oryzae* species complex include a number of weeds of rice, such as torpedo grass (*Panicum repens*) and cutgrass (*Leersia hexandra*), and cultivated grasses such as *S. italica* (Ou 1987). Most reproduction is asexual by mitotically produced spores (conidia) from lesions on above-ground plant parts. A single lesion can produce 2000–6000 conidia/day for up to 14 days, with multiple cycles of infection and reproduction during one growing season. Some conidia may disperse beyond the usual 1-m range (Ou 1987). The initial stages of infection are relatively nonspecific and could permit colonization of novel host species (Hamer *et al.* 1988; Peng and Shishiyama 1988; Heath *et al.* 1990, 1992). Although sexual reproduction occurs *in vitro* (isolates from Eleusine tend to show high fertility), the morphological structures, perithecia, are not observed in nature (Hebert 1971; Silué and Nottéghem 1990; Zeigler 1998). There is indirect evidence for genetic exchange and recombination in some rice-infecting populations (Valent *et al.* 1986; Kumar *et al.* 1999; Kato *et al.* 2000). Predominantly asexual reproduction would facilitate the emergence of pandemic clones specialized for pathogenicity against rice cultivars.

Overcoming both general and specific plant defenses potentially limits host-species shifts. Rice presents both partial and complete resistance toward *M. oryzae* (Marchetti 1983; Marchetti *et al.* 1987; Wang *et al.* 1989, 1994). Partial resistance reduces the overall disease severity and is usually controlled by multiple loci (Nottéghem *et al.* 1994). Complete resistance toward some genotypes of *M. oryzae* is conferred by major genes in rice (R-genes) that operate in a gene-for-gene manner

(Silué *et al.* 1992; Valent and Chumley 1994). Here, a plant resistance gene functions by recognizing pathogen genotypes carrying a corresponding avirulence (AVR) gene. Mutation or deletion in AVR genes permits the pathogen to evade detection by the plant, potentially allowing infection. A number of AVR genes recognized by rice R-genes have been identified in *M. oryzae* (Notteghem *et al.* 1994; Farman and Leong 1998; Orbach *et al.* 2000; Böhnert *et al.* 2004). R-genes are generally regarded to be determinants of host cultivar specificity, *i.e.*, resistance in some but not all rice varieties or cultivars (Ronald 1997; Jia *et al.* 2000). AVR/R-gene interactions, however, can also be determinants of host-species specificity, *e.g.*, virulence against *Eragrostis curvula* (weeping lovegrass) mediated by PWL2 (Sweigard *et al.* 1995). AVR genes associated with pathogen genotypes from hosts other than rice may be recognized by rice R-genes (Valent *et al.* 1991; Farman *et al.* 2002), preventing establishment on rice.

Consistent with the impediments to overcoming host defenses, the available molecular evidence suggests a single acquisition of pathogenicity on rice (Hamer *et al.* 1989; Borromeo *et al.* 1993; Shull and Hamer 1994). High copy numbers of the repetitive elements MGR586 (Pot3) and MGR583 characterize rice-infecting isolates (Hamer *et al.* 1989). Exceptions to this general pattern may represent additional host shifts with limited frequency and distribution (Leong *et al.* 1994). The clonal structure observed in rice-infecting populations (Levy *et al.* 1991; Xia *et al.* 1993; Zeigler 1998) supports the theory that contemporary clonal lineages derive from a single ancestral, sexual, phylogenetic lineage. Short of having in hand archeological samples of infected plants, which do not appear to exist, a chronology of host shifts and specialization in *M. oryzae* can best be inferred phylogenetically on the basis of contemporary samples.

We explored the origin of rice-infecting populations of *M. oryzae* with a multilocus phylogenetic analysis of 497 isolates of *M. oryzae* from culture collections and population samples. The origin was inferred for pandemic populations of *M. oryzae* established on rice, as well as the pattern of host shifts among closely related haplotypes from rice and other grasses. To test the hypotheses of one invasion of rice, parsimony, maximum-likelihood (ML), and Bayesian methods were implemented independently. From this, the timing of the host shift to rice was estimated. Several independent methods detected and localized conflict indicative of recombination. Detecting conflict was necessary (i) to implement statistical parsimony, (ii) to determine the combinability of loci, and (iii) to test the hypothesis of pandemic clonality associated with host shifts to rice. The phylogenetic framework was also used as a basis for assessing the role of AVR genes in determining host-species specificity and for determining the distribution of mating types. Pathogenicity tests were performed to assess host specificity in isolates representing phylogenetic lineages most closely related to rice-infecting populations.

MATERIALS AND METHODS

Fungal isolates: Isolates used in this study included both population samples and isolates from culture collections. Isolates collected by B. Couch were derived from single spores from single, sporulating lesions. Leaves containing lesions were incubated overnight in a moist chamber at room temperature to induce sporulation. Spores from a single lesion were streaked on water agar (Difco, Sparks, MD; 15 g agar/liter of distilled H_2O) and incubated overnight to allow spores to germinate. Single, germinating spores were subcultured on potato dextrose agar (PDA; Difco). Isolates were maintained on PDA. Storage of isolates was on filter paper inoculated and colonized by the fungus and then dried and stored in a desiccator at −20°.

DNA extraction: Isolates were grown in either 10 or 20 ml of liquid Fries medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g NH_4NO_3, 0.5 g $MgSO_4 \bullet 7\ H_2O$, 1.0 g KH_2PO_4, 0.1 g NaCl, 0.13 g $CaCl_2 \bullet 2H_2O$, 0.02 g $FeSO_4 \bullet 7H_2O$, 1.0 g yeast extract/liter of H_2O) on a rotary shaker for 3–5 days. Mycelium was harvested by filtration and dried under a vacuum. DNA was extracted from ~15 mg of dried mycelium following the procedure of Zolan and Pukkila (1986).

Marker development: Genomic regions containing DNA sequence polymorphisms were identified by direct sequencing of PCR products amplified from 19 reference isolates from five hosts: *Eleusine coracana* (goosegrass; 2 isolates), *Eragrostis curvula* (weeping lovegrass; 2), perennial ryegrass (2), rice (11), and Setaria millets (2). A total of 26 regions were screened. Eleven genomic regions containing portions of genes were amplified using previously published PCR primers: histone 3, histone 4, β-tubulin, aldehyde dehydrogenase, phosphate permease (Glass and Donaldson 1995), translation elongation factor 1 α, calmodulin, chitin synthase 1, actin, a ras protein, and the intergenic spacer region of the nuclear ribosomal DNA repeat (Carbone and Kohn 1999). Five genomic regions containing portions of genes were amplified using PCR primers designed from *M. grisea* sequences accessioned to GenBank: a xylanase gene (xyn22), a hydrophobin (MPG1), a nitrogen regulatory protein (NUT1), an ergosterol biosysnthesis gene (ERG2), and a G-protein α-subunit (MAGA). Eleven additional genomic regions were amplified using PCR primers designed from *M. grisea* genomic clones accessioned to GenBank that lacked significant homology to known genes. Primers designed in this study that amplify regions containing polymorphisms within the set of 21 reference isolates are in Table 1. Primers were designed with the aid of GENEWORKS (IntelliGenetics, Mountainview, CA). Polymorphic regions were localized to chromosomal location by performing BLASTN searches against the *M. grisea* genome sequence, assembly 2.1, produced by the Broad Institute, Massachusetts Institute of Technology (http://www.broad.mit.edu/annotation/fungi/magnaporthe).

Polymerase chain reaction and sequencing: PCR reactions were performed using either a Perkin-Elmer 9600 or a Perkin-Elmer 9700 thermocycler (Perkin-Elmer, Foster City, CA). PCR reactions contained 10 mm Tris-HCl (pH 8.3), 50 mm KCl, 200 μm of each dNTP, 0.5 μm of each oligonucleotide primer, 0.5 units AmpliTaq DNA polymerase (Roche Molecular Systems, Branchburg, NJ), 2.0 mm $MgCl_2$, and 10 μl of a 1:200 dilution of genomic DNA in a total volume of 20 μl. Thermocycling conditions were as follows: 95° for 8 min, 35 cycles of 95° for 15 sec, 55° for 20 sec, 72° for 60 sec, followed by 72° for 5 min and a 4° hold. PCR products were purified

TABLE 1

Polymerase chain reaction primers designed from *M. grisea sensu lato* genes or genomic clones

Locus name	Primer sequences 5′–3′	GenBank accession no. of sequence used for primer design
MPG1	AGAAGGTCGTCTCTTGCTGC; TTCACTCAACGCTGATCGC	L20685
NUT1	CATGATGCACGTCAATCTGC; CTGTGTCGGTGTCTGACGC	U60290
BAC6	ACATCATTGTCCTCCTCGTC; GTTCCTGTCATTCATTTTCAA	AQ114766
CH7-BAC7	AAGACACGAGAGCAAAGAAAGAAG; CGATACATTACAGTGCCTACGAA	AQ160231
CH7-BAC9	TGTAAGAAGCTCGGTGACTGAT; AGTGTTGCTTGAACGGCTAA	AQ398709

using QIAquick spin columns (QIAGEN, Mississauga, ON) following the manufacturer's instructions. Sequencing reactions were performed using an ABI Prism Big Dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA), following the manufacturer's instructions with the exception that only 25% of the recommended volume of terminator mix was used. Sequencing reactions were separated on an ABI 310 Genetic Analyzer. Electropherograms were interpreted with SEQUENCE ANALYSIS SOFTWARE v. 3.3 (Applied Biosystems). Forward and reverse sequences were imported into SEQUENCHER v. 4.1 (Gene Codes, Ann Arbor, MI), assembled into contigs, and visually checked for errors.

Single-strand conformation polymorphisms: Single-strand conformation polymorphisms (SSCPs) (Orita *et al.* 1989) were used to screen isolates for DNA sequence polymorphisms in 10 of the genomic regions identified as polymorphic: actin (ACT), BAC6, β-tubulin (βt-1), calmodulin (CAL), CH7-BAC7, CH7-BAC9, chitin synthase 1 (CHS), translation elongation factor 1α (EF-1α), MPG1, and NUT1. For SSCP screening, the βt-1 region was divided into two overlapping fragments. The 5′ region was amplified using the βt-1a primer (Glass and Donaldson 1995) and βt1-256R (5′-GGTCTGGATGTT GTTGGGGATCC-3′). The 3′ region was amplified using the βt-1b primer (Glass and Donaldson 1995) and βt1-271F (5′-TGAAGGAGGTCGAGGACCAG-3′). PCR reactions for SSCP analysis were performed in a 12-μl reaction volume using primer pairs end labeled with ^{33}P using T4 kinase (GIBCO BRL, Gaithersburg, MD), following manufacturer's instructions. Following thermocycling, 4 μl of stop dye [95% formamide, 10 mm EDTA (pH 8.0), 0.1% bromophenol blue, and 0.1% xylene cyanol] was added to each reaction and samples were heated to 95° for 8 min and chilled to 4°. One microliter of each sample was loaded on a 30 cm × 39 cm × 0.4 mm vertical polyacrylamide gel (6% polyacrylamide, 0.12% bisacrylamide, and 0.5 × tris-borate-EDTA). Gels were prepared with or without the addition of glycerol to a final concentration of 7%. Samples were separated at 150 V for 19–30 hr. Gels were blotted on filter paper and dried. Autoradiography was performed with Kodak BioMax film (Eastman Kodak, Rochester, NY) at room temperature. Three representatives of each allelic class were then sequenced to confirm their identity. For alleles with a frequency of less than three, all representatives were sequenced.

Phylogenetic analysis: Sequences for each haplotype were aligned using CLUSTAL X (Thompson *et al.* 1994) and edited manually using SEQUENCE ALIGNMENT EDITOR (Se-Al) v.1.0 α1 by Andrew Rambaut (http://evolve.zoo.ox.ac.uk/software.html). For coding indels, the complex indel-coding method of Simmons and Ochoterena (2000) was employed.

The basic units for phylogenetic analysis were multilocus haplotypes. Individuals sharing a haplotype had the same SSCP alleles for all 10 genomic regions screened in this study. Individuals were clustered into multilocus haplotypes using the neighbor-joining clustering method implemented in PAUP* v.4.0b10 for Macintosh (Swofford 2002). Since each SSCP allele was characterized by sequencing, each haplotype was represented by a unique multilocus DNA sequence.

Phylogenetic relationships among haplotypes were inferred using maximum-parsimony (MP) and ML methods implemented in PAUP* (Swofford 2002). Data from each genomic region were first analyzed alone, and then data for all regions were combined in a single analysis. For the MP analysis, heuristic searches were performed using starting trees obtained by stepwise addition, a simple addition sequence, and the branch-swapping algorithm of tree-bisection-reconnection. Visual inspection of branch-length tables, lists of character changes, and homoplasy matrices generated in PAUP* were used to locate parts of trees where there was phylogenetic conflict and to identify the haplotypes responsible for this conflict. Partitioned Bremer support values were calculated using the program TREEROT v. 2 (Sorenson 1999). For the PAUP tree searches specified by the TREEROT command file, 40 random addition replicates were performed for each search and the MaxTrees limit was set to 5000. Network reconstruction using statistical parsimony (Templeton *et al.* 1992) was performed using the program TCS v. 1.13 (Clement *et al.* 2000). For ML analysis, heuristic searches were performed using the Hasegawa, Kishino, and Yano (HKY) model of nucleotide substitution, identified using MODELTEST v. 3.06 (Posada and Crandall 1998) with empirically determined nucleotide frequencies and a molecular clock not enforced.

Bayesian analysis was performed using MRBAYES v. 3.0B3 (Huelsenbeck and Ronquist 2001). The settings used for the Metropolis Coupled Markov Chain Monte Carlo sampling were as follows: trees were sampled every 1000 generations from four chains run for 10,000,000 generations with a temperature setting for the heated chains of 0.2. The standard 4 × 4 model of DNA substitution was used. This gave an output of 10,000 trees, 999 of which were discarded as the "burn-in" for the Markov chain, leaving 9001 trees for estimation of the posterior probabilities for tree topologies.

Phylogenetic compatibility among polymorphic sites was visualized by generating a compatibility matrix using RETICULATE (Jakobsen and Easteal 1996). Split decomposition (Bandelt and Dress 1992) was used to visualize the tree-like and reticulate relationships among haplotypes. The program SPLITSTREE v. 2.4 (Dress *et al.* 1996) was used to construct a splits graph for the haplotypes. Iterations were performed until all haplotypes were resolved. The statistical support for each split was evaluated using bootstrapping (100 replicates).

Dating divergence: The assumption of a molecular clock was tested using the 1-d.f. method of Tajima (1993). Two indepen-

dent methods were used to estimate the divergence time for the host shift from Setaria millet to rice (Thorne *et al.* 1998; Sanderson 2002). First, in a Bayesian analysis the *a priori* estimate of the mean time between the tip and root of the tree was set to 10,000 years before present; according to the program manual, the standard deviation for this estimate is set equal to the mean (http://statgen.ncsu.edu/thorne/multidivtime.html). The burn-in value for the Markov chain was set at 100,000 cycles and divergence time estimates were based on 10,000 samples taken at intervals of 100 cycles from a single Markov chain. The burn-in was estimated by running the program with the burn-in set to zero, graphing node age estimates, and estimating the point at which the Markov chains converged on a "stable" value. In the second analysis, with the aid of r8s (http://ginger.ucdavis.edu/r8s/) the divergence time was estimated using the Langley-Fitch method with the truncated Newton optimization algorithm. Two separate runs were performed. In both runs the minimum age for the host shift from Setaria millet was set at 30 years BP (the age of our oldest sample from haplotype 10). For the first run, the root age was fixed at 10,000 years BP. In the second run, the root age was constrained to be between 10,000 and 40 years BP.

PCR test for presence or absence of AVR genes: The presence or absence of three AVR genes, *AVR-IRAT7* (*ACE1*), *AVR-Pita*, and *AVR-Co39*, was assessed using a plus/minus PCR test. A total of 59 isolates were screened. This group consisted of representatives from 28 haplotypes. The strains that were screened and their haplotype designations, indicated in parentheses, were: (1) G8, (2) G17, (3) VII-765, (3) 1152, (4) GF2-1, (4) G48, (5) Arcadia, (6) SJ5-1-2, (6) TT15-1, (7) SA1-1-1, (7) SA7-2-2, (8) SA2-1-1, (9) TT10-2, (10) BK19, (10) Guy11, (10) 75A36, (10) V-10, (11) 92A20, (11) 93M5, (11) 93T7, (11) A119, (12) BK6, (12) BK8, (12) ML39, (12) V33, (13) ML2, (13) ML7, (14) 75L23, (14) ML56, (14) ML91, (14) V-2, (15) Os/Or1-1-1, (15) Os/Or7-1-1, (15) Os/Or-30-1-1, (16) C9227-6, (16) DH1-2, (16) DH2-2, (16) DR1-2-1, (17) ML8, (18) DR42-3-2, (18) SA5-1-1, (19) GrF1, (19) McConnell, (20) 330, (20) 365, (21) Br81, (22) SAG00T12, (22) SAG00T6, (23) SAG00T3, (23) SAG00T4, (24) G199, (25) V-749, (25) VII-758, (26) V58, (26) V60, (27) VII-739, (27) VII953-1, (28) 1122, and (28) 94-115. For each isolate, PCR amplifications were performed using primers specific for each of the AVR genes. To ensure that the PCR primers amplified the expected product, representative amplicons were directly sequenced. For both *AVR-Pita* and *AVR1-Co39*, two independent amplifications were performed. As a positive control for DNA quality, amplifications were also performed using the primers for the actin region. Lack of amplification was interpreted as evidence for loss of function of the AVR gene due to deletion or rearrangement of the gene.

PCR primers were designed from GenBank sequences for *AVR-Pita* (AF207841) and *AVR1-Co39* (AF463528). Amplifications using the *AVR-Pita* primers *Pita*-F (5′-CGCCTTTTATTG GTTTAATTCG-3′) and *Pita*-R (5′-CCTCCATTCCAACACTA ACG-3′) were performed using an annealing temperature of 60°. Amplifications using the *AVR-Co39* primers *Co39*F-327 (5′-TGCGATATAATGGCCAAACA-3′) and *Co39*R-800 (5′-GACC GATCTGTCGGGAAGTA-3′) were performed using an annealing temperature of 55°.

The distribution of *ACE1* (Böhnert *et al.* 2004) was assessed using allele-specific PCR (Fudal 2004). The two major virulent *ACE1* genotypes (*ACE1-vir1*, *ACE1-vir2*) and the widely distributed *ACE1* avirulent allele (*ACE1-avr*) could be detected using two different PCR reactions, one for the virulent alleles and one for the avirulent allele, which were performed in replicate using the following primers: the *ACE1-avr* allele was specifically amplified using the primer pairs Guy11-X438+ (5′-GTTTC CGACATACTTTGCGCCC-3′) and Guy11-i1O- (5′-GAGCCGA CGTAGAGTTTTGGG-3′). The *ACE1-vir1* allele was specifically amplified using the primer pairs CM28-X438+ (5′-TTT CCCGACCTACTTTGCACCG-3′) and CM28-i1O- (5′-GACCCC ACGTACAGTTTGGCA-3′). The *ACE1-vir2* genotype was characterized by two PCR reactions performed with CM28 and Guy11 primers, respectively; the isolate has the ACE1 vir2 genotype if there is an amplicon for both reactions. Representative isolates with known *ACE1* genotypes (CM28, *ACE1-vir1*; PH14, *ACE1-vir2*; Guy11, *ACE1-avr*) were included as positive controls and water as a negative control in all replicates.

Mating-type frequencies among haplotypes: Clones in the plasmid vectors SK43 and SK44 of the mating-type idiomorphs (Kang *et al.* 1994) were purified using an alkaline lysis procedure (Sambrook *et al.* 1989) and were then digested with *Eco*RI and *Bgl*I following the manufacturer's instructions (Life Technologies, GIBCO BRL, Burlington, ON) to liberate the mating-type-specific sequences to be used as probes. These probe sequences were separated from plasmid DNA by gel electrophoresis and purified using a QIAquick gel extraction kit (QIAGEN) following the manufacturer's instructions. A 2.4-kb fragment from SK43 contained the MAT1-1-specific sequence, and a 2.3-kb fragment from SK44 contained the MAT1-2-specific sequence. Probes were labeled with ^{32}P using a nick translation kit (Invitrogen BV, Groningen, The Netherlands). Hybridization was performed at 60°. Autoradiography was performed with a Storm PhosphorImager (Amersham Bioscience, Piscataway, NJ).

For each isolate of *M. oryzae*, 10 μl of mini-prep DNA was digested with *Eco*RI (MBI Fermentas, Flamborough, ON) and separated on a 24-cm, 0.8% agarose gel for 45 min at 4 V/cm. Following electrophoresis, DNAs were blotted on GeneScreen Plus hybridization transfer membranes (Perkin-Elmer Life Sciences, Boston) following the manufacturer's instructions.

Pathogenicity testing: Two sets of pathogenicity tests were performed under standard inoculation conditions for this pathogen (Valent *et al.* 1991) or (Sallaud *et al.* 2003) by B. Valent (Kansas State University) and by D. Tharreau (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), respectively. In the first test, isolates assayed were representatives of haplotypes 6 (DR 22-1-1, isolated from *P. repens*), 10 (Guy 11 from rice), 18 (DR 52-2-1 from *L. hexandra*), and 19 (McConnell from *S. viridis*). A highly pathogenic field strain from China, O-137, was included as a positive control. Inoculations were performed on rice, variety YT16 (a highly susceptible variety), and on *P. repens* (accession no. 338659 from Morocco obtained from the USDA Plant Genetic Resources Conservation Unity, Griffin, GA). In the second test, isolates assayed were representatives of haplotypes 19 (Gr F1, Gr F2, Gr F3, Gr F4, and McConnell), haplotype 5 (Arcadia from *S. viridis*), haplotype 4 (Gr F2-1 from *S. faberii*), and haplotypes 18 (Dr 51-2-3 from *L. hexandra*). The rice-infecting isolate PH14 (PO6-6) was used as positive control. The rice cultivars, listed from most susceptible to most resistant, were Sariceltik, Maratelli, IR1529, and IR 64. Disease severity was scored on a scale from 0 to 5 on the basis of lesion types defined by Valent *et al.* (1991). Strains producing lesions of types 0 and 1, which do not sporulate, are considered nonpathogenic or avirulent. Strains producing lesions of type 2 are very weakly pathogenic although potentially capable of sporulating. Strains producing lesion types 3–5, which are capable of sporulating under conducive conditions, are considered pathogenic (Valent *et al.* 1991). Strains that produce type 5 lesions are considered highly pathogenic on the particular host genotype on which the reaction is observed. Some strains produce a variety of lesion types on a single plant. This reaction is termed "mesothetic" and scoring reflects the predominant lesion type.

TABLE 2

Summary of the frequency, host, and geographic association of the 37 multi-locus haplotypes identified in *M. oryzae*

Haplotype	Frequency	Host	Location
1	1	*E. curvula*	Japan
2	1	*E. curvula*	Japan
3	3	*Setaria* sp.	India
4	4	*Setaria* spp.	United States
5	1	*S. viridis*	United States
6	63	*P. repens*	Philippines and Vietnam
7	6	*P. repens*	Philippines
8	1	*P. repens*	Philippines
9	1	*O. sativa*	Vietnam
10	126	*O. sativa*[a]	[b]
11	10	*O. sativa*	United States
12	15	*O. sativa*	India
13	2	*O. sativa*	India
14	121	*O. sativa*	[c]
15	28	*O. sativa* × *O. rufipogon*	Philippines
16	45	*O. sativa*	Philippines and Vietnam
17	1	*O. sativa*	India
18	11	*L. hexandra*	Philippines and Vietnam
19	7	*S. viridis*	United States
20	2	*L. perenne*	United States
21	4	*T. aestivum*	Brazil
22	14	*S. secundatum*	United States
23	4	*S. secundatum*	United States
24	1	*E. coracana*	Rwanda
25	2	*Eleusine* sp.	India
26	1	*O. sativa*	India
27	2	*Eleusine* sp.	India
28	10	*Eleusine* spp.	India, China, Philippines, Ivory Coast
29	1	*O. sativa*	Russia
30	1	*E. indica*	Ivory Coast
31	1	*O. sativa*	France
32	1	*L. hexandra*	Ivory Coast
33	1	*L. hexandra*	Philippines
34	1	*O. sativa*	China
35	1	*B. mutica*	Philippines
36	2	*Eleusine indica*	Philippines
37	1	*E. indica*	Burkina Faso

The frequency, host of origin, and location of collections is given for each haplotype.

[a] Haplotype 10 is mainly associated with rice (123 isolates); 2 isolates were from barley (*H. vulgare*) and 1 isolate was from *P. clandestinum*. The 2 barley isolates from haplotype 10 are pathogenic to rice (D. Tharreau, unpublished results) and likely originated as a result of the infection of barley in North Thailand by rice-infecting isolates. Barley is a universal host plant for many nonrice, host-specific forms of blast.

[b] Members of haplotype 10 were from Burundi, China, Ivory Coast, French Guyana, India, Italy Japan, Korea, Philippines, Thailand, and the United States.

[c] Members of haplotype 14 were from Brazil, Cameroon, Ivory Coast, India, Japan, Madagascar, Philippines, Thailand, the United States, and Vietnam.

RESULTS

Marker development: Of the PCR primers tested, 10 amplified regions were polymorphic within the set of reference isolates: βt-1, EF-1α, CAL, CHS, ACT, a hydrophobin (MPG1), a nitrogen regulatory protein (NUT1), BAC6, CH7-BAC7, and CH7-BAC9. These regions are distributed throughout the genome and were localized to six of the seven *M. grisea* chromosomes as indicated in parentheses: CAL (II), CHS (III), EF-1α (III), ACT (IV), BAC6 (IV), βt-1 (VI), CH7-BAC7 (VII), CH7-BAC9 (VII), MPG1 (VII), and NUT1 (VII).

Haplotype identification and DNA sequence polymorphisms: A total of 37 unique multilocus haplotypes were identified by SSCPs and verified by sequencing in 497 isolates of *M. oryzae* (Table 2). A total of 83 base substitutions and 13 indels were identified in 3266 bp of aligned sequence. Two polymorphic sites, a C-G transversion in CH7-BAC7 and a 1-bp indel in EF1, could not be resolved using SSCPs and were excluded from all analyses. Of the 96 polymorphic sites, 63 were phylogenetically informative, *i.e.*, shared by two or more haplotypes. The other 33 sites were unique to one haplotype.

Parsimony analysis of haplotypes: Maximum-parsimony phylogenies inferred by analyzing data for each of the 10 loci separately contained very little homoplasy. Homoplasy occurs when a character state present in two taxa is not derived from a single common ancestor. In phylogenetic analyses of DNA sequences, homoplasious nucleotide sites are those in which identical mutations are inferred to have occurred multiple times. Single, most parsimonious, minimum-length trees were inferred using base substitutions from 7 of the 10 loci: ACT (2 steps), BAC6 (3 steps), BT1 (6 steps), CAL (5 steps), CH7-BAC7 (22 steps), CHS (2 steps), EF1 (4 steps), and NUT1 (3 steps). A single, most parsimonious tree of 24 steps, 1 step longer than the minimum-length tree, was inferred for the MPG1 region. The CH7-BAC9 region contained the most homoplasy. When parsimony analysis was conducted using only base substitutions, three most parsimonious trees of 14 steps were inferred, each 2 steps longer than the minimum-length tree. Two base positions were found to be homoplasious. These base positions are part of a group of three adjacent polymorphic nucleotide sites, positions 2112–2114 in the multiple sequence alignment, that may be a mutational hotspot subject to multiple hits.

Data from each locus analyzed individually provided very little phylogenetic resolution among haplotypes because many haplotypes shared identical sequences for some loci. To increase the resolution, data from all loci were combined in a single maximum-parsimony analysis. Using only base substitutions, 160 most parsimonious trees (MPTs) were inferred, each with a length of 118, 36 steps longer than the minimum-length tree (82 steps). Each tree had a consistency index (CI) of 0.695. One of the 160 MPTs is shown (Figure 1). Maximum-

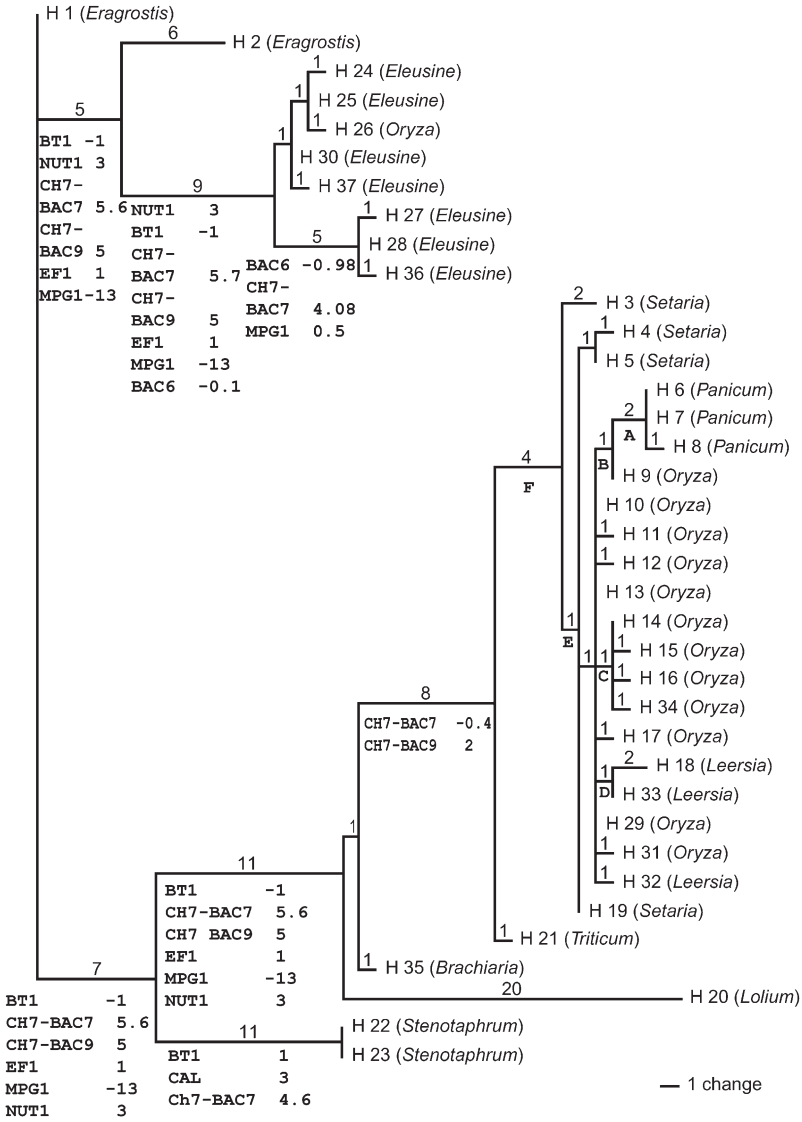

FIGURE 1.—One of 160 MPTs for the combined data from all 10 loci. Numerical designations prefixed by an "H" refer to a total of 37 unique multilocus haplotypes from 497 isolates of *M. oryzae*. The host genus or genera from which each haplotype was isolated is in parentheses. Only base substitutions were included in this analysis. Branch lengths are shown above each branch. Partitioned Bremer support values (BAKER and DESALLE 1997; BAKER *et al.* 1998) are given below long branches. Partitioned Bremer support values for branches indicated with letters are A: Bt1 (2), CH7-BAC7 (−0.9) and MPG1 (0.5). B: BAC6 (0.2), CH7-BAC7 (−1.6), CH7-BAC9 (−0.4), CHS (1.4), and MPG1 (0.9). C: BAC6 (0.3), CH7-BAC7 (−0.7), CH7-BAC9 (−0.5), CHS (0.5), and MPG1 (0.9). D: BAC6 (0.3), CH7-BAC7 (0.2), CH7-BAC9 (0.2), CHS (−0.2), and MPG1 (0.1). E: CH7-BAC7 (−0.9), CH7-BAC9 (−0.1), CHS (0.1), EF1 (1), and MPG1 (0.5). F: CH7-BAC7 (−0.4) and MPG1 (2).

likelihood analysis was also performed: maximum likelihood resolved identical branching to those parts of the parsimony tree that were free of phylogenetic conflict. When indels were included in the analysis, 80 MPTs were inferred. Each tree had a length of 142 steps, 44 steps longer than the minimum-length tree (98 steps). Each tree had a CI of 0.690. Additional trees were not recovered with heuristic searches performed with random addition of taxa in 10 replicates. Partitioned Bremer support indices were also calculated for each branch in the tree (Figure 1). The loci BT1, MPG1, and BAC6 have negative Bremer support values for a number of branches in this tree: BT1 at four branches, MPG1 at four branches, and BAC6 at two branches. Positive Bremer support values indicate that a branch is supported by data from a given locus and larger values are indicative of greater support. A negative value indicates that a branch is not supported by data from that locus. Larger negative values indicate greater conflict. Conflict among loci, as evidenced by branches with both positive and negative Bremer support values, was localized in most interior branches, as well as the bipartition containing haplotypes 2, 24–28, 30, 36, and 37. In contrast, in the bipartition-containing haplotypes 3–19, 29, and 31–34, there are only small negative Bremer support values, indicating that this part of the tree was relatively free of conflict. Conflict among specific loci was also apparent. BAC6 and CH7-BAC7 were in conflict at one node. Both BT1 and MPG1 were in conflict with four other loci: CH7-BAC7, CH7-BAC9, EF1, and NUT1.

To identify haplotypes responsible for phylogenetic conflict within the multilocus analysis, a homoplasy matrix was generated for the MP tree shown in Figure 1 (data not shown). There was very little conflict among haplotypes 3–19, 29, and 31–34, as evidenced by low or zero homoplasy values. In contrast, the other haplotypes

had higher homoplasy values in the range of 6–22, indicative of more conflict among these haplotypes, given this tree topology.

Phylogenetic conflict in this phylogeny was localized in the portions of the tree including haplotypes from *E. curvula*, *Eleusine* spp., *L. perenne*, *Brachiaria mutica*, and *Stenotaphrum secundatum*. The absence of conflict among loci permits the relationships among haplotypes 3–19, 29, and 31–34 from *O. sativa*, Setaria millets, *P. repens*, and *L. hexandra* to be represented by a tree. In contrast, relationships among the remaining haplotypes may reflect a history of recombination better represented as a reticulate network than as a tree.

Compatibility matrix analysis: Phylogenetic conflict is due either to recombination, in this case among phylogenetically informative sites, or to convergent evolution. A compatibility matrix was constructed to visualize patterns of phylogenetic compatibility among informative sites both within and among loci. Patterns of incompatibility consistent with recombination or convergent evolution were identified by visual inspection of the matrix. The compatibility matrix (Figure 2) is a visual representation of Hudson's four-gamete test performed on all pairwise combinations of sites (Hudson and Kaplan 1985). Incompatibility among sites indicates that they support conflicting bipartitions of taxa, with large blocks of incompatibility generally indicating recombination. As resolved in the phylogenetic analysis of each locus, the matrix showed that sites within a locus were generally compatible with the exception of the CH7-BAC9 and MPG1 regions, where there were large blocks of incompatible sites. Large blocks of incompatibility were detected among loci on the same chromosome as well as among loci on different chromosomes. The distances, in base pairs, between loci on the same chromosome were calculated using the complete genome sequence (Figure 2). Some distances could not be determined precisely due to gaps in the genome sequences: in these cases, minimum distances are given. In a second analysis (matrix not shown) performed on haplotypes 3–19, 29, and 31–34 from 21 polymorphic sites representing nine loci, only 2 sites were incompatible: site 1874 from the CH7-BAC7 region and site 2199 from the CHS region. Although seven of the nine loci contained multiple polymorphic sites, no blocks of incompatible sites were observed. This pattern of incompatibility is due either to a single recombination event between these loci, indicating low levels of recombination, or to the occurrence of multiple substitutions at 1 of the 2 sites.

Network analysis: The method of split decomposition (Bandelt and Dress 1992) was used to visualize the tree-like and non-tree-like relationships among haplotypes. A splits graph was constructed for the 37 *M. oryzae* haplotypes (Figure 3) using only base substitutions. As seen with the pattern of compatibility and homoplasy resolved in the parsimony analysis, a predominantly tree-like relationship among haplotypes 3–19, 29, and 31–34

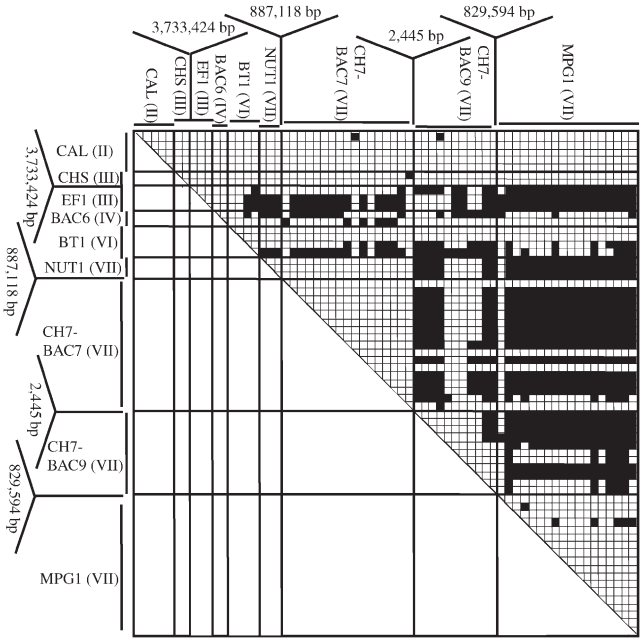

Figure 2.—A pairwise compatibility matrix for all phylogenetically informative sites from 9 of the 10 polymorphic loci generated using RETICULATE (Jakobsen and Easteal 1996). Incompatible sites that support conflicting phylogenies are indicated by solid boxes. Loci are labeled along the *x*- and *y*-axes. Thick bars bracketing the informative sites from each locus have been added as visual aids. As an additional visual aid, rows and columns containing blocks of sites from a single locus have been delimited using thick lines. Loci are organized according to chromosomal location and the chromosomal location of each locus is indicated with Roman numerals. For loci on a single chromosome, the minimum distance between loci (in base pairs) was calculated from the genome sequence and is indicated in the chevrons between loci. Exact distances between loci were available only for CH7-BAC7 and CH7-BAC9. The distances between all other loci are minimum estimates due to gaps in the genome sequence. The minimum distance reflects the distances between loci, ignoring gaps in the genome sequence.

was resolved. The reticulate relationships among the remaining haplotypes (1, 2, 20, 21–23, 24–28, 35, and 36) are depicted by multiple closed loops in the splits graph consistent with recombination.

Since a tree-like relationship could be inferred among haplotypes 3–19, 29, and 31–34, this phylogeny was converted to a network by means of statistical parsimony (Figure 4). Only base substitutions were included in the analysis. Haplotypes 10 and 14 from rice were the most frequent and widely distributed haplotypes in the sample. These two common, rice-infecting haplotypes belong to interior nodes in the network and have many mutational derivatives. Regardless of the root position for the network, the common, widespread, rice-infecting haplotypes appear to have a single origin. Midpoint rooting was used to polarize this network and determine the direction of host shifts. The root was placed at the midpoint of the longest branch in the phylogeny (the mid-

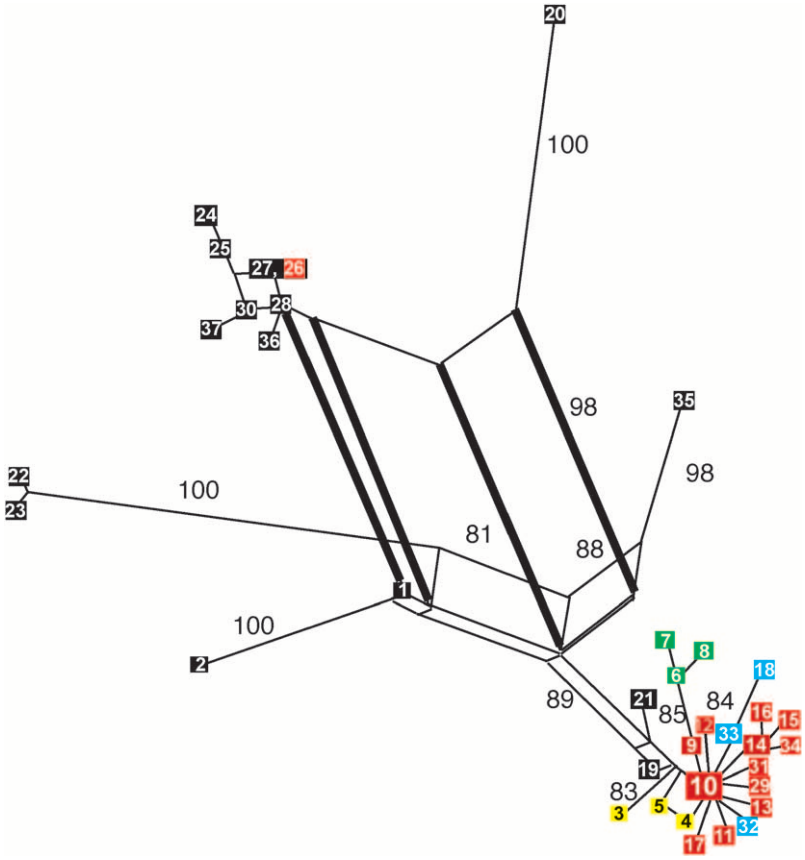

FIGURE 3.—A splits graph for *M. oryzae* haplotypes based on combined data from all 10 loci. Numerical designations refer to each of a total of 37 multilocus haplotypes from 497 isolates of *M. oryzae*. Indels were excluded from the analysis. Nodes containing sampled haplotypes are indicated by boxes. Colors are used to indicate the host of origin for each haplotype: red, *O. sativa*; yellow, Setaria; blue, *L. hexandra*; green, *P. repens*; black, other grass hosts. Bootstrap support values are indicated along one edge of each split; all edges of a given split have the same bootstrap support. For splits represented by bands of parallel edges the bootstrap value is indicated only along one edge. In reticulate parts of the network, splits are represented by bands of parallel edges. For example, the split that partitions the haplotypes into the two sets {1, 2, 3–19, 21–23, 29, 31–35} and {20, 24–28, 30, 36, 37} is represented by the four parallel edges indicated by thick lines.

point root). Since we are representing infraspecific relationships in our network, we assume that haplotypes that occur at tips of the network, such as H8, are more recently derived than internal haplotypes, such as H10. Midpoint rooting was the only reasonable way of rooting this network and places the root at an internal, ancestral node within the network. Using another *M. oryzae* haplotype for rooting was impossible, given the reticulate relationship among isolates from different hosts and the arbitrariness of choosing one haplotype as an outgroup. The most closely related species, *M. grisea*, was too divergent for use as an outgroup. Coalescent rooting was not reliable, given the differences in sampling among hosts. Our conclusion of a single host shift to rice is not affected by the placement of the root at any other internal or tip node. Placement of the root at different internal nodes, however, would alter the inferred pattern of host shifts. This method placed the root between haplotypes 3 and 19. Given this root position, the pattern of host shifts in *M. oryzae* is from Setaria millets to rice and from rice to *L. hexandra* and *P. repens*. The phylogenetic conflict between sites from CH7-BAC9 and CHS that was resolved in the compatibility analysis is represented in this network as a single closed loop.

Bayesian analysis: Bayesian analyses corroborated the branching order resolved in the network analysis. A Bayesian phylogeny was constructed from the combined data for 10 loci for all haplotypes (Figure 5). The main features of this phylogeny are strong support (P = 1.000) for the clade-containing haplotypes 3–19, 29, and 31–34. This clade contains isolates from Setaria millets, *L. hexandra*, *P. repens*, and most of the rice isolates. If we assume that the root position falls between haplotypes 3 and 19, then haplotypes from Setaria millets are basal to the clade-containing haplotypes from rice, *L. hexandra*, and *P. repens* (P = 0.710). Haplotypes from both *L. hexandra* and *P. repens* are derived from rice-infecting haplotypes (P = 1.000). A single Setaria millet-infecting haplotype (H4) also appears to be derived from a rice-infecting haplotype. This analysis suggests the following pattern of host shifts: Setaria millet to rice, followed by shifts from rice to *L. hexandra* and *P. repens*.

Dating divergence of the rice-infecting lineage: Two events in the history of rice cultivation that may have facilitated a host shift to rice were domestication [~7000 years ago; (CRAWFORD and SHEN 1998)] and the Green Revolution (35 years ago). We attempted to determine which event was most likely associated with the origin of rice-infecting populations by dating the divergence of rice-infecting haplotypes. If molecular evolution in the lineage containing rice-infecting isolates is clock like, the number of base substitutions separating two haplotypes should be proportional to their divergence times. A molecular evolutionary clock hypothesis could

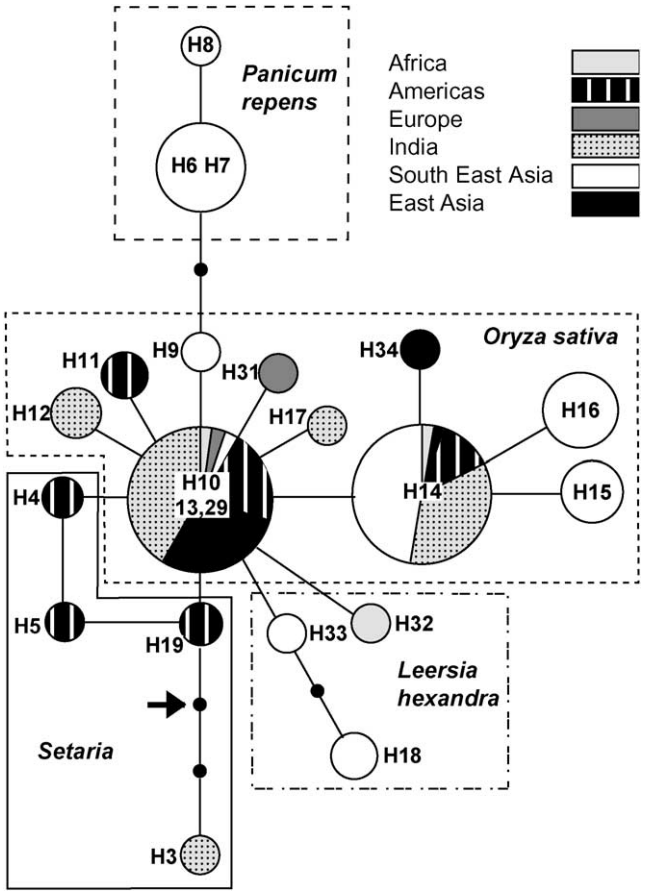

FIGURE 4.—The statistical parsimony network constructed for haplotypes 3–19, 29, and 31–34. Each haplotype or group of haplotypes is represented by a solid circle. The relative frequency of each haplotype is linearly related to the area of each circle. The geographic locations where isolates originated are represented by patterns indicated in the legend. Groups of haplotypes isolated from the same host are indicated with boxes and the host of origin is indicated within each box. Although haplotype 10 is mainly associated with rice (123 isolates), 2 isolates were from barley (*Hordeum vulgare*) and 1 isolate was from *Pennisetum clandestinum*. Each branch represents a single mutational step. In cases where haplotypes are connected by branches of more than a single mutational step, unsampled intermediates differing by a single mutational step have been inferred and are indicated by small solid circles. The midpoint root is indicated by an arrow. Haplotypes 13 and 29 are distinguished from haplotype 10 only by indels, not nucleotide polymorphisms; indels were not employed in this analysis.

not be rejected for isolates from rice, Panicum, and Leersia ($P \geq 0.75$); therefore, estimates of the timing of divergence of rice-infecting isolates were generated using rate smoothing. Calibration of the molecular clock to estimate divergence times in years from the number of substitutions separating two haplotypes requires an estimate of the age for at least one node in a phylogeny. For the Bayesian analysis we used an *a priori* estimate of 10,000 years BP for the age of the node separating haplotypes from Setaria millet from haplotypes from rice, *L. hexandra*, and *P. repens*. In this analysis the standard deviation of the estimate was set to 10,000 years BP. This time estimate was chosen as a maximum age for this node because it predates rice domestication, but still falls within the time frame of use of rice by humans (12,000 BP). Using the method of SANDERSON (2002) implemented in the program r8s, two estimates for the age of the node separating haplotypes from Setaria millet from haplotypes from rice, *L. hexandra*, and *P. repens* were used; the age was fixed at 10,000 BP, and the age was constrained to be between 10,000 and 40 BP. Both methods for estimating divergence time gave comparable results. The Bayesian method of THORNE *et al.* (1998) resolved a divergence time of 7291 BP (95% C.I. = 26,042–453 BP). Using the method of SANDERSON (2002), a divergence time of 5373 BP was estimated when the root age was fixed at 10,000 BP. When the root age was allowed to vary between 10,000 and 40 BP, the divergence time was estimated at 2574 BP. Both dating methods were in agreement in supporting an earlier divergence time, rather than a later time associated with the Green Revolution.

Distribution of AVR genes: The distribution of *AVR-Pita* and *AVR1-Co39* was evaluated using a simple plus/minus PCR test (Figure 5). Failure to amplify a gene was taken as evidence for the loss of gene function as a result of complete or partial deletion or rearrangement. To address the issue of false negatives due to low-quality DNA, PCR amplifications of the actin region were performed on all DNA dilutions. The actin region was amplified in all isolates, indicating that DNA quality was sufficient for PCR amplification. It is possible, however, that mutation in or around the primer-binding site could also have contributed to false negatives. The distribution of *ACE1* was evaluated using allele-specific PCR. A single loss of function for *AVR-Co39* is associated with the major Oryza-infecting lineage, excluding haplotypes 9 and 26. In contrast, the loss of *AVR1-Pita* is not associated with a single clade on any host. The two *ACE1* virulent genotypes (*vir1* and *vir2*) were detected only in haplotype H14 and its derived haplotype H16 (Figure 4), suggesting a common origin for both virulent genotypes. H16 has only *vir1*, consistent with an origin of this virulence genotype in a subpopulation of H14.

Pathogenicity testing: In all pathogenicity tests the rice-infecting controls were generally highly pathogenic toward rice (Table 3). Nonrice isolates were generally avirulent or weakly pathogenic toward rice. Some isolates did produce lesions on rice cultivars but these lesions were of a less severe type or fewer in number than those produced by the rice-infecting controls. The isolate Dr 51-2-3 from cutgrass produced type 3 lesions on the variety IR1529-380-3 and a few type 3 lesions on Sariceltik. The isolates Gr F2-1, Arcadia, and Gr F4 from *S. viridis* produced some lesions on the more susceptible cultivars Maratelli and Sariceltik. Isolates from *S. viridis* and *P. repens* were not pathogenic toward rice and pro-

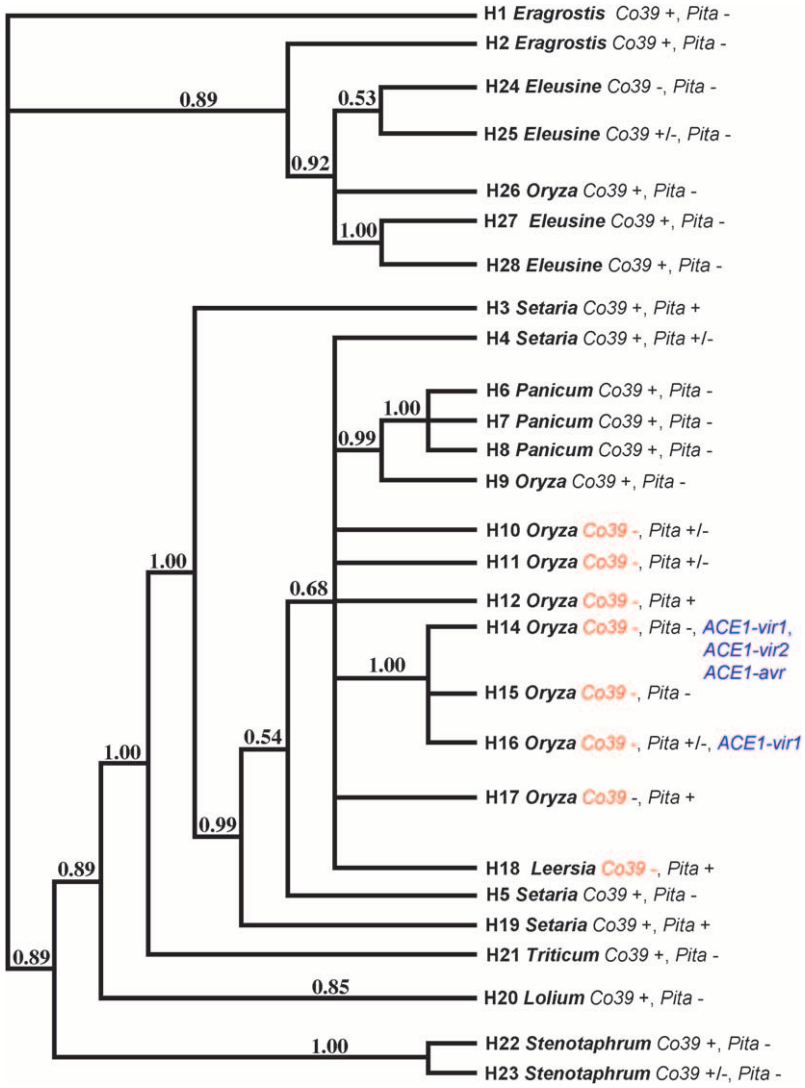

FIGURE 5.—The Bayesian consensus tree for *M. oryzae* haplotypes inferred from the combined data from all 10 loci. Numerical designations prefixed by an "H" refer to a total of 37 unique multilocus haplotypes from 497 isolates of *M. oryzae*. The values above the branches are the posterior probabilities associated with the branch supporting a particular clade. Branch lengths are proportional to the posterior probability for each branch. Posterior probabilities are not given for branches terminating in a single haplotype. Haplotypes 13 and 29 are not shown since they are distinguished from haplotype 10 only by indels, not nucleotide polymorphisms; indels were not employed in this analysis. The host of origin for each haplotype is given to the right of the haplotype designation. The presence or absence of *AVR1-Co39* and *AVR-Pita* were assessed using a PCR test: "+" indicates that an amplification product was produced, "−" indicates that no amplification product was produced and that presumably the gene is absent from these strains. For the *ACE1* gene, two haplotypes had alleles that confer a virulent phenotype, *ACE1-vir1*, *ACE1-vir2*. All other rice-infecting isolates are likely to be ACE1-avr, although some rare virulent isolates with PCR amplicons similar to that of ACE1-avr cannot be detected with current methods.

duced some lesions of the type capable of sporulating on *P. repens*. Even though strain O-137 was highly pathogenic toward rice, it was only weakly pathogenic on *P. repens*.

Mating-type testing: Both MAT1-1 and MAT1-2 were identified within rice-infecting populations of *M. oryzae* with MAT1-2 being the most frequent (Table 4). The common, widely distributed haplotypes, 10 and 14, were composed of isolates with either MAT1-1 or MAT1-2. Haplotypes derived from 10 and 14 (Figure 4) included isolates with only one mating type. Only single mating types were identified within populations from torpedo grass (MAT1-2) and cutgrass (MAT1-1).

DISCUSSION

Contrary to the expectation that rice should be a large target for multiple, continuing, parasite host shifts, we resolved a single, relatively old origin from Setaria millet for a pandemic lineage of *M. oryzae* on rice. In turn, rice crops have been a source for invasions of the common rice field weeds, cutgrass and torpedo grass. Nearly all rice-infecting isolates fell within a single phylogenetic lineage, as would be expected if the contemporary, globally distributed, rice-infecting populations of *M. oryzae* originated as a result of a single host shift to rice. The branching phylogenetic relationship of rice-infecting haplotypes is evidence of clonal amplification of this lineage following invasion of rice. Reticulate relationships indicating recombination were not observed among these haplotypes. Isolates from torpedo grass and cutgrass compose two distinct lineages derived from a rice-infecting ancestor. The host specificity associated with these lineages appears to be maintained by differences in pathogenicity among isolates from different hosts. Isolates from hosts other than rice either were not pathogenic toward rice or were much less pathogenic when compared with typical rice-infecting isolates (Table 3). Although serial culturing could reduce pathogenicity, control isolates, also serially cultured, demon-

TABLE 3

Pathogenicity testing of *M. oryzae* isolates from *O. sativa*, *S. italica*, *L. hexandra*, and *P. repens*

Fungal strain (haplotype)	Source	*O. sativa* varieties					*P. repens*
		IR64	IR1529-380-3	Maratelli	YT 16	Sariceltik	
O-137	*O. sativa*	2–3	NT	NT	5	5	Some type 2 on edges of older leaves
Guy 11 (10)	*O. sativa*	2	2	5	4 mesothetic[a]	5	0 (1 type 1)
Ph 14 (=PO6-6)	*O. sativa*	5	5	5	NT	5	NT
DR 52-2-1 (18)	*L. hexandra*	NT	NT	NT	Sparse type 2	NT	0
Dr 51-2-3 (18)	*L. hexandra*	0	3	0	NT	Some type 3	NT
DR 22-1-1 (6)	*P. repens*	NT	NT	NT	0	NT	Few lesions on some plants
Gr F2-1 (4)	*S. viridis*	0	0	0 (2 type 5)	NT	0	NT
Arcadia (5)	*S. viridis*	3	0	0 (2 type 4)	NT	Some type 5	NT
McConnell (19)	*S. viridis*	0	0	0	0	0	Few lesions on some plants
Gr F1 (19)	*S. viridis*	0	0	0	NT	0	NT
Gr F2 (19)	*S. viridis*	0	0	0	NT	0	NT
Gr F3 (19)	*S. viridis*	0	0	0	NT	0	NT
Gr F4 (19)	*S. viridis*	0	0	0	NT	2–3 (1 type 4)	NT

Rice varieties are arranged from left to right in order of increasing susceptibility. Lesion types are as follows: type 0, no signs of infection (no lesions or necrotic spots); type 1, small brown lesions up to 0.5 mm in diameter and lacking visible tan centers; type 2, small (~1 mm in diameter) lesions with tan centers and dark brown margins; type 3, small (~2 mm in length) diamond-shaped lesions with tan centers and dark brown margins; type 4, intermediate (~3–4 mm in length) diamond-shaped lesions; type 5, large diamond-shaped lesions (VALENT *et al.* 1991). Strains producing lesions of types 0 and 1, which do not sporulate, are considered nonpathogenic or avirulent. Strains producing lesions of type 2 are very weakly pathogenic although potentially capable of sporulating. Types 3–5 are able to sporulate under conducive conditions and are considered pathogenic. NT indicates that an isolate was not tested on the host.

[a] A mesothetic reaction is one in which a variety of lesion types are produced on a single plant and scoring reflects the predominant lesion type.

strated expected pathogenicity. Among several possible mechanisms for host specialization, the pathogen avirulence determinant *AVR-Co39* was widespread in haplotypes from hosts other than rice, potentially preventing infection of rice cultivars equipped with the corresponding resistance gene *Pi-Co39*.

A single origin of a pandemic lineage of *M. oryzae* on rice could represent (i) one invasion of rice, (ii) multi-

TABLE 4

Distribution and frequency of mating-type idiomorphs among *M. oryzae* haplotypes from *O. sativa*, *P. repens*, and *L. hexandra*

Haplotype	Host	No. of isolates screened	Frequency of mating-type idiomorphs (%)		
			MAT 1-1	MAT1-2	Unknown
6	*P. repens*	63	0	82.5	17.5
7	*P. repens*	6	0	33.3	66.7
8	*P. repens*	1	0	100	0
9	*O. sativa*	1	0	100	0
10	*O. sativa*	112	56.3	42.9	0.8
11	*O. sativa*	9	0	88.8	11.2
12	*O. sativa*	14	0	100	0
13	*O. sativa*	2	0	100	0
14	*O. sativa*	112	7.1	89.3	3.6
15	*O. sativa*	28	0	100	0
16	*O. sativa*	45	0	97.8	2.2
17	*O. sativa*	1	100	0	0
18	*L. hexandra*	15	100	0	0
32	*L. hexandra*	1	100	0	0

ple invasions with the descendants of one invasion surviving genetic drift, or (iii) multiple invasions mostly undetected in our sampling due to their limited geographical distribution. During the course of this study the sample was increased three times in increments of ~100 isolates without resolving additional host jumps, making alternative (iii) unlikely. Sampling from diverse locations should have resolved additional host shifts where new genotypes have become established on rice even if they have not become globally distributed. A potential bias in this sample is that very weakly pathogenic isolates would be underrepresented since only plants with visible lesions, capable of sporulating, were sampled. It is possible that changes in deployment of rice cultivars and agronomic practices post-Green Revolution resulted in a bottleneck consistent with alternative (ii), reducing genetic diversity in the pathogen and obliterating traces of pathogen lineages other than the one rice-infecting lineage detected in this study. This explanation would not undermine a time estimate of lineage divergence predating the Green Revolution.

The observed direction of host shifts in *M. oryzae* is especially interesting, considering that both millet and rice were domesticated in China and appear to have partially co-occurred early in the history of agriculture in Asia. Italian millet is the domesticated species, with a primary gene pool comprising two weedy Setaria millets, *S. viridis* and *S. faberi*, which are recognized by some as distinct species (de Wet *et al.* 1979; Dekker 2003). Although the widespread contact between Italian millet and rice in China occurred only recently and within written historical time, there is some archeological evidence for earlier co-occurrence of these two crops (Chang 1986; Crawford 1992; Higman and Lu 1998).

The sequence of host shifts was robust even with alternative root positions, and the relationship of haplotypes in the lineage including millet, rice, cutgrass, and torpedo grass was not obscured by phylogenetic conflict. Midpoint rooting resolved a single root position between haplotypes 3 and 19 (Figure 4). The relationships among these haplotypes are well supported by parsimony, network, and Bayesian methods (Figures 1, 3, 4, and 5). The only conflict in this part of the phylogeny is represented by the closed loop containing haplotypes 4, 5, 10, and 19, due to one base substitution (Figure 4). A single origin for rice-infecting haplotypes is still resolved regardless of the cause of this conflict, either recombination or multiple hits.

In addition to the single dominant lineage of haplotypes associated with rice, a lineage was resolved comprising several haplotypes from goosegrass (Eleusine), but also including one haplotype (26) associated with rice (Figures 1, 3, and 5). Haplotype 26 was sampled at low frequency, as 2 of 112 isolates from India, and was not recovered in subsequent sampling from Southeast Asia. Haplotype 26 may represent a decaying chain of infection in which the parasite has crossed a species boundary but cannot sustain a positive reproductive rate (May *et al.* 2001).

The pandemic spread of a single HIV subtype suggests that the accidents of genetic drift have dominated the population structure on a global scale, resulting in a single widely distributed lineage (Sharp *et al.* 1994; Chen *et al.* 1997; Hutchinson 2001). If the pattern seen in HIV is a model for *M. oryzae*, additional host shifts to rice may have evaded detection because haplotypes recently established on rice could have limited dispersal. If remnants of additional host shifts exist, finding them will require regional sampling without predictive data for selecting localities. The association between early domestication of rice and Setaria millet suggests that the host shift to rice occurred in China, where evidence for additional shifts may yet be found.

In the statistical parsimony analysis we included only base substitutions. Indels were excluded from the analysis because they are not accommodated by the substitution model (Templeton *et al.* 1992). When indels were added to the network inferred from statistical parsimony, more haplotypes associated with rice were distinguished and a single 4-bp indel appeared to distinguish rice-infecting haplotypes (excluding haplotypes 9 and 26) from all other haplotypes. Inclusion of this character necessitates inferring an unsampled common ancestor (with no host association) for the haplotypes from rice, cutgrass, and torpedo grass. In contrast, in all analyses in which only base substitutions were used, haplotypes 10, 13, and 29 were clearly resolved as ancestral to haplotypes from cutgrass and torpedo grass. These haplotypes show nearly complete association with rice; haplotype 13 consists of 2 isolates from rice, haplotype 29 consists of 1 isolate from rice, and haplotype 10 consists of 126 isolates, 123 from rice. In haplotype 10, 2 of the 3 isolates from hosts other than rice were from barley, but were pathogenic on rice, suggesting that they represented invasions of barley from rice (D. Tharreau, unpublished results). We propose the most parsimonious explanation that both this indel and the high copy number of Pot3 evolved in rice-infecting populations following the host shifts from rice to torpedo grass and cutgrass. We are assuming that an unsampled ancestor of H10 had a low copy number of *Pot*3 and lacked the indel. This haplotype gave rise to the cutgrass and torpedo grass populations. Following host shifts to cutgrass and torpedo grass, a small deletion occurred in the CAL locus and the copy number of *Pot*3 increased.

Amplification of Pot3 also appears to have occurred in the rice-infecting lineage following host shifts from rice to cutgrass and torpedo grass. In the present study, isolates from rice have a high copy number of this element while isolates from cutgrass and torpedo grass have a much lower copy number (data not shown). Isolates from Setaria millets have been reported to show variable hybridization to Pot3, with some isolates showing weak or no hybridization (Farman 2002) and others

showing a medium copy number in Setaria millet isolates but lower than that in rice isolates (Eto *et al.* 2001; Kang *et al.* 2001; Tosa *et al.* 2004). An independent amplification of Pot3 in some Pennisetum isolates with low genetic similarity to isolates from Setaria millets and rice has also been reported (Kang *et al.* 2001). These observations suggest that amplification of transposable elements may occur frequently and, on a contemporary timescale, in populations on a variety of hosts. The high copy number of this element in all populations from rice supports a single origin for rice-infecting populations and also a history of asexual reproduction. Prolonged asexual reproduction in ascomycetous fungi favors the amplification of repetitive elements since transposon inactivation requires a sexual cycle (Cambareri *et al.* 1989), although sexual reproduction may stimulate transposition of some elements such as MAGGY (Eto *et al.* 2001).

Multilocus analysis of *M. oryzae* by means of parsimony, network, and Bayesian methods all resolved a pattern of clonality in haplotypes from rice and their close relatives from Setaria millets, cutgrass, and torpedo grass and a pattern of recombination among haplotypes from other hosts. This was done by sequential analyses, first on data from each locus and then on combined data from all loci, which were used to identify the cause of homoplasy, to identify haplotypes responsible for homoplasy, and to localize homoplasy to specific regions within the phylogenies inferred for all haplotypes.

Patterns of incompatibility among polymorphic sites from all loci, represented in a compatibility matrix, indicate that recombination is likely to be responsible for the majority of the homoplasy observed in these data (Carbone and Kohn 2001). Convergent evolution of SSCP alleles is not an important source of homoplasy in this study; no cases of convergent evolution were identified by direct sequencing of SSCP alleles and comparable levels of conflict among loci were resolved when the analysis was restricted to isolates for which all loci were sequenced (data not shown). Sites are considered incompatible if two nucleotides at different sites support conflicting phylogenetic topologies. Incompatibility among sites is due either to recombination among sites or to convergent evolution resulting from multiple substitutions. Several arguments support recombination in this data set. First, to develop the large blocks of incompatible sites identified among the loci CH7-BAC7, CH7-BAC9, and MPG1 in the compatibility analysis (Figure 2), it is unlikely that the alternative to recombination, multiple hits, would occur repeatedly in adjoining nucleotide positions, given an infinite-sites or nearly infinite-sites model of DNA sequence evolution (Kimura 1969; Hudson and Kaplan 1985). On the other hand, in comparisons of other pairs of loci with fewer polymorphic sites, it is difficult to distinguish recombination from multiple hits as the cause of incompatibility. But given that multiple hits are expected to be rare, the presence of clusters of incompatible sites, as in this data set, is more likely due to recombination. Further, the majority of these loci with fewer polymorphisms occur on different chromosomes among which segregation should be random. The small numbers of substitutions in these loci do not suggest that they fall within mutational hotspots, in which multiple hits would be expected to occur with higher frequency. Finally, no polymorphic sites that had more than two bases segregating within the sample were identified, consistent with an infinite-sites model of DNA sequence evolution (Kimura 1969).

The complementary analyses using parsimony, network, and Bayesian methods resolved patterns of recombination consistent with the compatibility matrix. Predominantly tree-like relationships, expected with clonality (Carbone *et al.* 1999), were inferred among haplotypes from wheat, rice, Setaria millets, cutgrass, and torpedo grass. Reticulate relationships, expected with recombination, were required to infer relationships among the remaining haplotypes from other hosts. The resolution of tree-like relationships among rice-infecting haplotypes is consistent with previous reports based on other types of data that support that *M. oryzae* has a predominantly clonal population structure and is known to reproduce only asexually (Ou 1987; Levy *et al.* 1993; Park *et al.* 2003).

We detected negligible, ambiguous signals of recombination in the rice-Setaria millet-cutgrass-torpedo grass lineage. Two sites were found to be incompatible, due to either recombination or convergent evolution. If recombination is the cause of this incompatibility, it appears to have occurred prior to the divergence of haplotypes from rice, cutgrass, and torpedo grass (Figure 4). The detection of both mating types in each of two haplotypes, MAT1-1 and MAT1-2 in haplotypes 10 and 14, is further support of the hypothesis of historical but not contemporary recombination. Although both mating types are present in some populations, this in itself is not indicative of contemporary sexual reproduction. In the absence of selection or drift, both mating types could be maintained. Since no incompatibility was detected among phylogenetically informative sites in these lineages, the presence of both mating types is not strong evidence for active sexual reproduction.

The main difficulty in detecting recombination is the low power of the four-gamete test (Hudson and Kaplan 1985). The probability of detecting recombination increases with the number of segregating sites on either side of the recombination breakpoint and the recombination rate between regions (Hudson and Kaplan 1985). We chose multiple markers located on different chromosomes or distributed along the same chromosome to increase the number of potentially segregating sites screened, the number of intervals in which recombination can occur, and the rate of recombination among loci, given that the recombination rate should be maxi-

mal among unlinked loci. The incompatibility among closely linked loci on the same chromosome suggests that the recombination detected is due to true sexual reproduction rather than parasexuality, which results most frequently in recombination among chromosomes (Pontecorvo 1956). Recombination within some rice-infecting populations has been reported, as evidenced by allelic associations in pairwise comparisons, and multilocus variance analysis of DNA fingerprints with the multilocus probe MGR586 (Pot3) and single-locus probes (Zeigler *et al.* 1997; Zeigler 1998). However, the best evidence for recombination within rice-infecting populations would be recovery of perithecia from naturally infected hosts in the field, with ascospores showing segregation with respect to parental types. To date, reports of female sexual fertility and high genetic diversity in pathogen populations in southern China and Thailand seem to be the best evidence available of at least local sexual reproduction against a background of predominant clonality (Hayashi *et al.* 1997; Mekwatanakarn *et al.* 1999).

A history of recombination has probably obscured the origins of host-specific populations of *M. oryzae* on perennial ryegrass and wheat. A possible strategy is to identify closely related populations on other hosts, possibly tall fescue (*Festuca arundinacea*) for perennial ryegrass (Urashima *et al.* 1999; Viji *et al.* 2001; Farman 2002; Tosa *et al.* 2004), among which tree-like patterns of relationship may exist.

In addition to revealing patterns of recombination, the multilocus approach detected divergence among isolates from different hosts not previously detected with a single locus. For example, wheat and goosegrass (Eleusine) isolates cannot be distinguished solely on the basis of the intergenic transcribed spacer of the nuclear rDNA repeat, but are distinguished on the basis of some RFLP markers, including the FOSBURY retrotransposon specific to Eleusine (Viji *et al.* 2001; Tosa *et al.* 2004). In the present study, isolates from the two hosts are distinguished as a number of host-specific, multilocus haplotypes whose phylogenetic relationships are obscured by recombination.

Host shifts to rice from other grasses may be limited by either AVR-R interactions preventing initiation of a host shift or the subsequent inability to become established on rice. The presence of *AVR1-Co39* in nonrice haplotypes and its absence in haplotypes from rice corroborate the earlier findings of Farman (2002). The presence of *AVR1-Co39* in haplotypes 6, 7, 8, and 9 suggests that this gene was lost in the rice-infecting lineage following the host shift to topedo grass but does not necessarily imply that the loss of this gene is essential for rice parasitism. Although the rice R-gene that recognizes *AVR1-Co39* has been defeated by the pathogen, it may still constrain additional host shifts to rice. Following the initial stages of the host shift to rice, lack of fitness of the invader on the new host could impede establishment. Pathogenicity tests support this argument; in our greenhouse test, nonrice isolates were generally not pathogenic toward rice. The avirulent or weakly pathogenic reactions toward rice observed for haplotypes from Setaria millet, torpedo grass, and cutgrass suggest that they are poorly adapted to rice, and although they may be able to establish some level of infection *in vitro*, they may not be able to become established on rice.

Changes in the virulence spectra of rice-infecting isolates also occur within rice-infecting populations. Breakdown of complete resistance, conferred by major resistance genes, often occurs rapidly in the field in response to the release of new, blast-resistant cultivars (Notteghem *et al.* 1994). Our results suggest that gain of virulence toward rice varieties with the resistance gene corresponding to *ACE1* has occurred in a single rice-infecting haplotype. The two *ACE1* virulent genotypes (*ACE1*-vir1 and *ACE1*-vir2) were detected only in isolates from haplotype 14, suggesting a common origin for both virulent genotypes. Haplotype 16, descended from haplotype 14, has only *ACE1*-vir1, suggesting that it was derived from a subpopulation of haplotype 14, carrying the *ACE1*-vir1 allele. The amplification of transposable elements such as Pot3 in the genomes of rice-infecting isolates and the resultant genetic instability could result in deletion of AVR genes and in the rapid evolution of virulence toward newly deployed R-genes.

Given that *M. oryzae* on rice seems to have a single origin and has been moved around the world in association with rice cultivation, what genetic determinants are unique to the rice-infecting lineage? What mechanisms in rice could potentially prevent the establishment of new *M. oryzae* genotypes? Our understanding of the genetic determinants conveying virulence and pathogenicity in the rice-infecting lineage of *M. oryzae* will deepen by considering the origins and dynamics of these determinants implicit in the evolutionary relationships among haplotypes from other hosts. Presumably the lineages of Setaria millets, torpedo grass, and cutgrass are genetically very similar to the lineage from rice since they are closely related. They are, however, expected to lack some or all of the genetic determinants responsible for high fitness on rice. Representatives of these haplotypes would be prime candidates as parents in crosses to identify the genetic determinants of fitness on rice common to all rice-infecting haplotypes.

We thank C. Vera-Cruz and E. Borromeo for their help with sampling in the Philippines as well as M. Marchetti and all those who contributed isolates to this study. For technical support we thank Caroline Sirjusingh and Andrea Chan at the University of Toronto, Henri Adreit and Joelle Milazzo at Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), and Melinda Dalby at Kansas State University. This work was supported by a Discovery Research Grant from the Natural Sciences and Engineering Research Council of Canada to L. M. Kohn and by scholarships to B. Couch from the Ontario Graduate Scholarship program. Part of this work was supported by the European Community RESIDIV ICA4-CT2000-30021 project grants to the CIRAD and Centre National

de la Recherche Scientifique teams. A United States Department of Agriculture-Cooperative State Research and Extension Service grant to M. Milgroom (L. M. Kohn coapplicant) supported the development of markers and fieldwork of B. Couch.

LITERATURE CITED

Antonovics, J., M. Hood and J. Partain, 2002 The ecology and genetics of a host-shift: *Microbotryum* as a model system. Am. Nat. **160:** S40–S53.

Baker, R. H., and R. DeSalle, 1997 Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. **46:** 654–673.

Baker, R. H., X. Yu and R. DeSalle, 1998 Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. **9:** 427–436.

Bandelt, H. J., and A. W. Dress, 1992 Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1:** 242–252.

Böhnert, H. U., I. Fudal, W. Dioh, D. Tharreau, J.-L. Notteghem *et al.*, 2004 A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. Plant Cell **16:** 2499–2513.

Borromeo, E. S., R. J. Nelson, J. M. Bonman and H. Leung, 1993 Genetic differentiation among isolates of *Pyricularia* infecting rice weed hosts. Phytopathology **83:** 393–399.

Brasier, C. M., and K. W. Buck, 2001 Rapid evolutionary changes in a globally invading fungal pathogen (Dutch elm disease). Biol. Invasions **3:** 223–233.

Bremer, K., 2002 Gondwanan evolution of the grass alliance of families (Poales). Evolution **56:** 1374–1387.

Cambareri, E. B., B. C. Jensen, E. Schabtach and E. U. Selker, 1989 Repeat-induced G-C to A-T mutations in *Neurospora*. Science **244:** 1571–1575.

Carbone, I., and L. M. Kohn, 1999 A method for designing primer sets for speciation studies in filamentous ascomycetes. Mycologia **91:** 553–556.

Carbone, I., and L. M. Kohn, 2001 A microbial population–species interface: nested cladistic and coalescent inference with multilocus data. Mol. Ecol. **10:** 947–964.

Carbone, I., J. B. Anderson and L. M. Kohn, 1999 Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. Evolution **53:** 11–21.

Chang, K. C., 1986 *The Archaeology of Ancient China*. Yale University Press, New Haven, CT.

Chang, T. T., 1976 The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica **25:** 425–441.

Chen, Z., A. Luckay, D. L. Sodora, P. Telfer, P. Reed *et al.*, 1997 Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. J. Virol. **71:** 3953–3960.

Cheng, C., R. Motohashi, S. Tsuchimoto, Y. Fukuta, H. Ohtsubo *et al.*, 2003 Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol. Biol. Evol. **20:** 67–75.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1660.

Couch, B. C., and L. M. Kohn, 2002 A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea*. Mycologia **94:** 683–693.

Crawford, G. W., 1992 Prehistoric plant domestication in East Asia, pp. 7–38 in *The Origins of Agriculture: An International Perspective*, edited by C. W. Cowan and P. J. Watson. Smithsonian Institution Press, Washington, DC.

Crawford, G. W., and C. Shen, 1998 The origins of rice agriculture: recent progress in East Asia. Antiquity **72:** 858–866.

Crepet, W., and G. D. Feldman, 1991 The earliest remains of grasses in the fossil record. Am. J. Bot. **78:** 1010–1014.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Dekker, J., 2003 The foxtail (*Setaria*) species-group. Weed Sci. **51:** 641–656.

de Wet, J. M. J., L. L. Oestry-Stidd and J. I. Cubero, 1979 Origins and evolution of foxtail millets (*Setaria italica*). J. Agric. Trop. Bot. Appl. **26:** 53–64.

Dress, A., D. Huson and V. Moulton, 1996 Analyzing and visualizing sequence and distance data using SplitsTree. Discrete Appl. Math. **71:** 95–109.

Eto, Y., K. Ikeda, I. Chuma, T. Kataoka, S. Kuroda *et al.*, 2001 Comparative analyses of the distribution of various transposable elements in *Pyricularia* and their activity during and after the sexual cycle. Mol. Gen. Genet. **264:** 565–577.

Farman, M. L., 2002 *Pyricularia grisea* isolates causing gray leaf spot on perennial ryegrass (*Lolium perenne*) in the United States: relationship to *P. grisea* isolates from other host plants. Phytopathology **92:** 245–254.

Farman, M. L., and S. A. Leong, 1998 Chromosome walking to the *AVR1-CO39* avirulence gene of *Magnaporthe grisea*: discrepancy between the physical and genetic maps. Genetics **150:** 1049–1058.

Farman, M. L., Y. Eto, T. Nakao, Y. Tosa, H. Nakayashiki *et al.*, 2002 Analysis of the structure of the *AVR1–CO39* avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. Mol. Plant Microbe Interact. **15:** 6–16.

Fry, J. D., 1996 The evolution of host specialization: Are trade-offs overrated? Am. Nat. **148:** S84–S107.

Fudal, I., 2004 Etude du gene d'avirulence ACE1 de *Magnaporthe grisea*, agent pathogène du riz: analyse de l'expression du gène ACE1 et évolution dans les populations de *Magnaporthe grisea*. Ph.D. Thesis, University of Orsay, Orsay, France.

Glass, N. L., and G. C. Donaldson, 1995 Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous Ascomycetes. Appl. Environ. Microbiol. **61:** 1323–1330.

Grass Phylogeny Working Group, 2001 Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Mo. Bot. Gar. **88:** 373–457.

Hamer, J. E., R. J. Howard, F. G. Chumley and B. Valent, 1988 A mechanism for surface attachment in spores of a plant pathogenic fungus. Science **239:** 288–290.

Hamer, J. E., L. Farrall, M. J. Orbach, B. Valent and F. G. Chumley, 1989 Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. Proc. Natl. Acad. Sci. USA **86:** 9981–9985.

Hayashi, N., C. Li, J. Li, M. Iwano, H. Naito *et al.*, 1997 Distribution of fertile *Magnaporthe grisea* fungus pathogenic to rice in Yunnan Province, China. Ann. Phytopathol. Soc. Jpn. **63:** 316–323.

Heath, M. C., R. J. Howard, B. Valent and F. G. Chumley, 1990 Interactions of two strains of *Magnaporthe grisea* with rice, goosegrass, and weeping lovegrass. Can. J. Bot. **68:** 1627–1637.

Heath, M. C., R. J. Howard, B. Valent and F. G. Chumley, 1992 Ultrastructural interactions of one strain of *Magnaporthe grisea* with goosegrass and weeping lovegrass. Can. J. Bot. **70:** 779–787.

Hebert, T. T., 1971 The perfect stage of *Pyricularia grisea*. Phytopathology **61:** 83–87.

Higman, C., and T. L.-D. Lu, 1998 The origins and dispersal of rice cultivation. Antiquity **72:** 867–877.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Huelsenbeck, J. P., and F. Ronquist, 2001 MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754–755.

Hutchinson, J. F., 2001 The biology and evolution of HIV. Annu. Rev. Anthropol. **30:** 85–108.

Jakobsen, I. B., and S. Easteal, 1996 A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Comput. Appl. Biosci. **12:** 291–295.

Jia, Y., S. A. McAdams, G. T. Bryan, H. P. Hershey and B. Valent, 2000 Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. **19:** 4004–4014.

Kang, S., F. G. Chumley and B. Valent, 1994 Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. Genetics **138:** 289–296.

Kang, S., M. H. Lebrun, L. Farrall and B. Valent, 2001 Gain of virulence by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. Mol. Plant Microbe Interact. **14:** 671–674.

Kato, H., M. Yamamoto, T. Yamaguchi-Ozaki, H. Kadouchi, Y. Iwamoto *et al.*, 2000 Pathogenicity, mating ability and DNA

restriction fragment length polymorphisms of Pyricularia populations isolated from Gramineae, Bambusideae and Zingiberaceae plants. J. Gen. Plant Pathol. **66** (1): 30–47.

Khush, G. S., 2001 Green revolution: the way forward. Nat. Rev. Genet. **2:** 815–822.

Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61:** 893–903.

Kumar, J., R. J. Nelson and R. S. Zeigler, 1999 Population structure and dynamics of *Magnaporthe grisea* in the Indian Himalayas. Genetics **152:** 971–984.

Landschoot, P. J., and B. F. Hoyland, 1992 Gray leaf spot of perennial ryegrass turf in Pennsylvania. Plant Dis. **76:** 1280–1282.

Leong, S., M. Farman, J. Smith, A. Budde, Y. Tosa *et al.*, 1994 Molecular genetic approach to the study of cultivar specificity in the rice blast fungus, pp. 87–110 in *Rice Blast Disease*, edited by R. Zeigler, S. Leong and P. Teng. CAB International, Wallingford, Oxfordshire, UK.

Levy, M., J. Romao, M. A. Marchetti and J. E. Hamer, 1991 DNA fingerprinting with a dispersed repeated sequence resolves pathotype diversity in the rice blast fungus. Plant Cell **3:** 95–102.

Levy, M., F. J. Correa Victoria, R. S. Zeigler, S. Xu and J. E. Hamer, 1993 Genetic diversity of the rice blast fungus in a disease nursery in Colombia. Phytopathology **83:** 1427–1433.

Long, D. H., J. C. Correll, F. N. Lee and D. O. TeBeest, 2001 Rice blast epidemics initiated by infested rice grain on the soil surface. Plant Dis. **85:** 612–616.

Lu, B.-R., K. L. Zheng, H. R. Qian and J. Y. Zhang, 2002 Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor. Appl. Genet. **106:** 101–106.

MacLeod, A., A. Tait and C. M. R. Turner, 2001 The population genetics of *Trypanosoma brucei* and the origin of human infectivity. Philos. Trans. R. Soc. Lond. B Biol. Sci. **356:** 1035–1044.

Marchetti, M., 1983 Dilatory resistance to rice blast in USA rice. Phytopathology **73:** 645–649.

Marchetti, M., L. Xinghua and C. Bollich, 1987 Inheritance of resistance to *Pyricularia oryzae* in rice cultivars grown in the United States. Phytopathology **77:** 799–804.

May, R. M., S. Gupta and A. R. McLean, 2001 Infectious disease dynamics: What characterizes a successful invader? Philos. Trans. R. Soc. Lond. B Biol. Sci. **356:** 901–910.

Maynard Smith, J., N. H. Smith, M. O'Rourke and B. G. Spratt, 1993 How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:** 4384–4388.

Mekwatanakarn, P., W. Kositratana, T. Promraksa and R. Zeigler, 1999 Sexually fertile *Magnaporthe grisea* rice pathogens in Thailand. Plant Dis. **83:** 939–943.

Nottéghem, J. L., D. Tharreau, D. Silué and E. Roumen, 1994 Present knowledge of rice resistance and strategies for *Magnaporthe grisea* pathogenicity and avirulence gene analysis, pp. 155–165 in *Rice Blast Disease*, edited by R. Zeigler, S. Leong and P. Teng. CAB International, Wallingford, Oxfordshire, UK.

Oh, H. S., Y. Tosa, N. Takabayashi, S. Nakagawa, R. Tomita *et al.*, 2002 Characterization of an *Avena* isolate of *Magnaporthe grisea* and identification of a locus conditioning its specificity on oat. Can. J. Bot. **80:** 1088–1095.

Oka, H. I., 1988 *Origin of Cultivated Rice*. Elsevier, Amsterdam.

Orbach, M. J., L. Farrell, J. A. Sweigard, F. G. Chumley and B. Valent, 2000 A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. Plant Cell **12:** 2019–2032.

Orita, M., Y. Suzuki, T. Sekiya and K. Hayashi, 1989 Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics **5:** 874–879.

Ou, S. H., 1987 *Rice Diseases*. The Commonwealth Mycological Institute, Kew, Surrey, UK.

Park, S. Y., M. G. Milgroom, S. S. Han, S. Kang and Y. H. Lee, 2003 Diversity of pathotypes and DNA fingerprint haplotypes in populations of *Magnaporthe grisea* in Korea over two decades. Phytopathology **93:** 1378–1385.

Peng, Y. L., and J. Shishiyama, 1988 Temporal sequence of cytological events in rice leaves infected with *Pyricularia oryzae*. Can. J. Bot. **66:** 730–735.

Pontecorvo, G., 1956 The parasexual cycle. Annu. Rev. Microbiol. **10:** 393–400.

Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Ronald, P. C., 1997 The molecular basis of disease resistance in rice. Plant Mol. Biol. **35:** 179–186.

Sallaud, C., M. Lorieux, E. Roumen, D. Tharreau, R. Berruyer *et al.*, 2003 Identification of five new blast resistance genes in the highly blast-resistant rice variety IR64 using a QTL mapping strategy. Theor. Appl. Genet. **106:** 794–803.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanderson, M. J., 2002 Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. **19:** 101–109.

Second, G., 1982 Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. **57:** 25–57.

Sharp, P. M., D. L. Robertson, F. Gao and B. H. Hahn, 1994 Origins and diversity of human immunodeficiency viruses. AIDS **8:** S27–S42.

Shull, V., and J. Hamer, 1994 Genome structure and variability in *Pyricularia grisea*, pp. 65–86 in *Rice Blast Disease*, edited by R. Zeigler, S. Leong and P. Teng. CAB International, Wallingford, Oxfordshire, UK.

Silué, D., and J. L. Nottéghem, 1990 Production of perithecia of *Magnaporthe grisea* on rice plants. Mycol. Res. **94:** 1151–1152.

Silué, D., J. L. Nottéghem and D. Tharreau, 1992 Evidence of a gene-for-gene relationship in the *Oryza sativa-Magnaporthe grisea* pathosystem. Phytopathology **82:** 577–580.

Simmons, M. P., and H. Ochoterena, 2000 Gaps as characters in sequence-based phylogenetic analysis. Syst. Biol. **49:** 369–381.

Sorenson, M. D., 1999 *TreeRot*, Version 2. Boston University, Boston.

Sweigard, J. A., A. M. Carroll, S. Kang, L. Farrall, F. G. Chumley *et al.*, 1995 Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. Plant Cell **7:** 1221–1233.

Swofford, D. L., 2002 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0. Sinauer Associates, Sunderland, MA.

Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

Talbot, N. J., 2003 On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. Annu. Rev. Microbiol. **57:** 177–202.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Thorne, J. L., H. Kishino and I. S. Painter, 1998 Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. **15:** 1647–1657.

Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, S. F. Brenière *et al.*, 1991 Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc. Natl. Acad. Sci. USA **88:** 5129–5133.

Tosa, Y., K. Hirata, H. Tamba, S. Nakagawa, I. Chuma *et al.*, 2004 Genetic constitution and pathogenicity of *Lolium* isolates of *Magnaporthe oryzae* in comparison with host species-specific pathotypes of the blast fungus. Phytopathology **94:** 454–462.

Urashima, A. S., Y. Hashimoto, L. D. Don, M. Kusaba, Y. Tosa *et al.*, 1999 Molecular analysis of the wheat blast population in Brazil with a homolog of retrotransposon MGR583. Ann. Phytopathol. Soc. Jpn. **65:** 429–436.

Valent, B., 1990 Rice blast as a model system for plant pathology. Phytopathology **80:** 33–36.

Valent, B., and F. G. Chumley, 1994 Avirulence genes and mechanisms of genetic instability in the rice blast fungus, pp. 111–134 in *Rice Blast Disease*, edited by R. S. Zeigler, S. A. Leong and P. S. Teng. CAB International, Wallingford, Oxfordshire, UK.

Valent, B., M. S. Crawford, C. G. Weaver and F. G. Chumley,

1986 Genetic studies of fertility and pathogenicity in *Magnaporthe grisea* (*Pyricularia oryzae*). Iowa State J. Res. **60:** 559–594.

Valent, B., L. Farrall and F. G. Chumley, 1991 *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. Genetics **127:** 87–101.

Vaughan, D. A., and H. Morishima, 2003 Biosystematics of the genus *Oryza*, pp. 27–65 in *Rice: Origin, History, Technology, and Production*, edited by C. W. Smith and R. H. Dilday. John Wiley & Sons, Hoboken, NJ.

Viji, G., B. Wu, S. Kang and W. Uddin, 2001 *Pyricularia grisea* causing gray leaf spot of perennial ryegrass turf: population structure and host specificity. Plant Dis. **85:** 817–826.

Wang, G., D. Mackill, M. Bonman, S. McCouch, M. Champoux *et al.*, 1994 RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics **136:** 1421–1434.

Wang, Z., D. J. Mackill and J. M. Bonman, 1989 Inheritance of partial resistance to blast in *indica* rice cultivars. Crop. Sci. **29:** 848–853.

Whitlock, M. C., 1996 The red queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. Am. Nat. **148:** S65–S77.

Xia, J. Q., J. C. Correll, F. N. Lee, M. A. Marchetti and D. D. Rhoads, 1993 DNA fingerprinting to examine microgeographic variation in the *Magnaporthe grisea* (*Pyricularia grisea*) population in two rice fields in Arkansas. Phytopathology **83:** 1029–1035.

Yamanaka, S., I. Nakamura, H. Nakai and Y.-I. Sato, 2003 Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. Genet. Res. Crop Evol. **50:** 529–538.

Zeigler, R. S., 1998 Recombination in *Magnaporthe grisea*. Annu. Rev. Phytopathol. **36:** 249–275.

Zeigler, R. S., R. P. Scott, H. Leung, A. A. Bordeos, J. Kumar *et al.*, 1997 Evidence of parasexual exchange of DNA in the rice blast fungus challenges its exclusive clonality. Phytopathology **87:** 284–294.

Zolan, M. E., and P. J. Pukkila, 1986 Inheritance of DNA methylation in *Coprinus cinereus*. Mol. Cell. Biol. **6:** 195–200.

Communicating editor: P. J. Oefner